Combining Synthetic Controls and VARs:

On the Estimation of Causal Effects in Time Series Data

Jörg Breitung[a], **Lennart Boldwin**[b], **Justus Töhns**[b]

[a]University of Cologne, Chair of Statistics and Econometrics
[b]University of Cologne, Chair of Statistics and Econometrics
Supervised by Prof. Jörg Breitung

Abstract

We argue that applications of Synthetic Control (SC) are faced with a self-selection problem. That is, the method is primarily applied to non-complex data structures that are straightforward to forecast, given the availability of donors in the post-treatment period. Using simulation studies, we show that the high interpretability of SC comes at the cost of poor predictions and forecasts, which are especially pronounced if the data generating process contains a time series structure. To address this issue, we introduce the intricacy-statistics that informs the applied researcher whether or not the data at hand exceeds a level of time series structure that SC can handle. If the case, more flexible methodologies that combine the strengths of SC and conventional time series techniques promise more accurate predictions and forecasts. Hence we introduce the new Vector Autoregressive Synthetic Control (VARSC) estimator, that takes in account both the time series structure and the availability of donors. In order to implement these ideas, we introduce the R-package varsc that provides ready-to-use functions to compute the intricacy-statistics and, based on the magnitude of the statistics, the functionalities to estimate either the SC or the VARSC model. To probe the performance of our methodology outside the experimental setting, we apply it to three existing applications of SC: Specifically, we show that our proposed model performs equally well like the SC-method. The result is striking because in contrast to the SC-model, our models gets along without the informational contend of potential covariates.

Keywords: *Synthetic Control; Causality; VAR*

List of Acronyms

ADH Abadie, Diamond, and Hainmueller

GDP Gross Domestic Product

SC Synthetic Control

USA United States of America

VAR Vector Autoregression

VARSC Vector Autoregressive Synthetic Control

Contents

1. Introduction

SC is cool. But Vector Autoregression (VAR) also.

Method was introduced by ADH. Make clear that this label comes from [Doudchenko and Imbens, 2016]

2. Literature Review 2-3 pages

What must be clear by now

- ...

2.1. Synthetic Control

The SC method was developed by Alberto Abadie and colleagues in a series of influential papers ([Abadie and Gardeazabal, 2003], [Abadie et al., 2010], [Abadie et al., 2015]). The method is designed to estimate the causal effect of a treatment in a setting with a single treatment unit and a number of potential control units. Pre- and post-treatment data are observed for the treatment and control units for the outcome of interest as well as for a set of covariates. The SC-procedure combines aspects of the matching and difference-in-difference literature and can therefore be interpreted as a relative of the causal inference literature introduced by [Rubin, 1974]. Similar to many other microeconometric methods, the objective is to distinguish causation from correlation and to assess the magnitude and significance of treatments in observational case studies.
In their canonical 2003 article, Abadie and Gardeazabal evaluate the causal economic effects of conflict using terrorist conflicts in the Basque Country as a comparative case study. In their specific application example, they find that terrorist conflicts caused the per capita Gross Domestic Product (GDP) of the treatment unit (Basque Country) to decline by about 10% relative to the synthesized control unit.

some more words on other findis and things they did The next appropriate setting for an application of the SC method was the introduction of a large-scale tobacco control program implemented in the state of California in the United States of America (USA) in 1988.

2.2. Overview

[Abadie, 2021] read.
[Athey and Imbens, 2016] read.

2.3. Application

[Born et al., 2019] read.
[Cho, 2020] read.
[Cunningham, 2021] read.
[Funke et al., 2020] read.

2.4. Methodological Background

[Hainmueller et al., 2011] read.
[Abadie and Imbens, 2006] not read.

[Abadie and Imbens, 2002] not read.
[Doudchenko and Imbens, 2016] read.
[Ferman, 2021] read.
[Frangakis and Rubin, 2002] not read.
[Rosenbaum and Rubin, 1983] not read
[Rubin, 1974] not read.

2.5. Extensions/ Developments

[Abadie and L'Hour, 2021] read.
[Amjad et al., 2018] read.
[Ben-Michael et al., 2021] read.
[Ben-Michael et al., 2021] not read.
[Kellogg et al., 2021] not read.
[Kuosmanen et al., 2021] not read.
[Muhlbach and Nielsen, 2019] read.

Developments
[Arkhangelsky et al., 2021] not read
[Athey et al., 2017] not read.
[Brodersen et al., 2015] read.
[von Brzeski et al., 2015] read.
[Hartford et al., 2017] read.

2.6. Testing

[Andrews, 2003] not read.
[Cattaneo et al., 2021] not read.
[Chernozhukov et al., 2019] not read.
[Chernozhukov et al., 2021] not read.
[Firpo and Possebom, 2018] not read.
[Hahn and Shi, 2017] read.

2.7. Time Series Econometrics

[Martin et al., 2012] read.
[Harvey and Thiele, 2020] read.
[Breitung and Knüppel, 2021] partially read.

3. Theory

What must be clear by now

- **Consider case without covariates**
- **Make clear that SC is a weighted average of the donors**
- **...**

In this chapter, we propose an alternative SC-estimator to assess the magnitude of treatment effects in observational settings. To establish a general basis, let us first describe the contextual environment of the estimation. Similar to the setting as introduced by ADH, we consider a framework with $J+1$ panel units indexed by $j = 0, 1, ..., J$ that are observed over a time horizon of T periods. Without loss of generality, assume that unit $j = 0$, is exposed to the treatment at period $t = T_0$ with $1 < T_0 < T$ and that there are no treatment anticipation and contamination. To contextualize these assumptions, [Abadie et al., 2010] argue that in the presence of anticipation effects, T_0 could be shifted back in time until the assumptions seems plausible. If panel units in the donor pool[1] are affected by the treatment (contamination) as it is likely in Brexit-application, those units could be removed from the sample prior to the estimation. Our goal is to evaluate the causal effect of the treatment, the specific functional form of which has yet to be specified.

The following theoretical argumentation is structured as follows: To have a common denominator, we first describe the canonical estimation procedure as proposed by ADH. Next we build intuition by considering a very simple static scenario with only two donor units and one treatment unit. We then generalize this idea to the case with many potential donors. The main difference to the setting of ADH is that we remove some of the weight constraints and that we analyze a situation without covariates. The former distinction guides us to the field of regularization in order to prevent our method from overfitting. The latter drastically reduces the data requirements but causes our algorithm to estimate the counterfactual with a significantly smaller information set. This fact leads us to our main contribution: The integration of multivariate time series approaches into the SC-algorithm.

3.1. ADH Case

We start by presenting the SC-method in its original form as introduced ADH. Besides introducing the general estimation technique, we also want to elaborate on the proposed hypothesis testing procedure of ADH. For the sake of comparability and due to its notational and inhat clarity, we borrow the employed notation.

[1] To ensure direct comparability with the SC literature, we adopt most of the commonly used terms. For example, control group units are labeled as 'donors'.

Setup

The estimation task can be constituted by the potential outcome framework as introduced by [Neyman, 1923] and refined by [Rubin, 1974]. Let $Y_{j,t}^I$ be the (potential) outcome for unit j at point t in the presence of the intervention. Likewise, let $Y_{j,t}^N$ be the (potential) outcome for j at point t in the absence of the intervention. Similar to ADH, we define the treatment effect as

$$\delta_{j,t} = Y_{j,t}^I - Y_{j,t}^N$$

and introduce the indicator variable $D_{j,t}$ that takes on the value 1 if unit j is treated at period t and the value 0 otherwise. Given the assumed absence of anticipation and contamination, we observe the following outcome

$$Y_{j,t} + D_{j,t}\delta_{j,t} = \begin{cases} Y_{j,t}^N & (\text{if } j = 0 \text{ and } t < T_0) \text{ or } j \geq 1, \\ Y_{j,t}^N + \delta_{j,t} & \text{if } j = 0 \text{ and } t \geq T_0 \end{cases}$$

The goal to estimate the causal treatment effect $(\delta_{0,T_0}, ..., \delta_{0,T})$ therefore boils down to the estimation of the counterfactuals of each unit $j = 0$ in the post-treatment phase $(Y_{0,T_0}, ..., Y_{0,T})$. The basic idea of ADH is to estimate these counterfactuals as a weighted average of the donor outcomes, using a data-driven approach to compute the weights. Intuitively, the weights are computed such that they optimally predict the outcomes and a set of explanatory variables of the treatment unit in the pre-intervention phase, conditional on having a percentage interpretation. To operationalize this intuition, let $\boldsymbol{Y_j} = (Y_{j,1}, ..., Y_{j,T_0})'$ be the vector of observed outcomes in the pre-treatment phase for unit j[2]. To distinguish treatment unit and donors, ADH denote the $(T_0 \times 1)$-vector for the treatment unit as $\boldsymbol{Y_1}$ and the $(T_0 \times J)$-vector for the donors as $\boldsymbol{Y_0}$. Moreover, a set of K covariates is observed for all panel units for $t = 1, 2, ..., T$, yet only the pre-treatment values are needed for the weight-calculation. Therefore, let $\boldsymbol{X_1}$ denote the $(K \times 1)$-vector of covariates for $\boldsymbol{Y_1}$ and let $\boldsymbol{X_0}$ denote the $(K \times J)$-matrix of explanatory variables for $\boldsymbol{Y_0}$.

Hypothesis Testing

lets go

3.2. Simple Static Extension

Consider a very simple framework for analyzing the causal effect of a treatment for unit $i = 0$ and two units in the control group $i = 1, 2$. It is assumed that before the intervention at time period $t = T_0$ the units have a joint distribution of the form

$$\boldsymbol{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ before } T_0.$$

[2] For instance, in the canonical example of [Abadie and Gardeazabal, 2003], $\boldsymbol{Y_j}$ would be the vector of GDPs.

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)'$ and $\boldsymbol{\Sigma}$ is some positive definite covariance matrix with Choleski decomposition $\boldsymbol{\Sigma} = \boldsymbol{R}\boldsymbol{R}'$ and $\boldsymbol{R}$ is an *upper* triangular matrix. Assume that the intervention affects the mean of the first variable such that $\mathbb{E}(Y_0) = \mu_0 + \delta$ after the intervention, whereas the means of the other two variables remain unaffected. Accordingly, δ represents the treatment effect on Y_0.

We are interested in deriving an optimal estimator for the counterfactual

$$\widehat{Y}_0^N = \mathbb{E}(Y_0|\delta = 0, Y_1, Y_2) \text{ after } T_0.$$

Let $Q = R^{-1}$ and q denotes the first row of Q, then

$$q'\boldsymbol{y} = q'\boldsymbol{\mu} + \epsilon,$$

where $\epsilon \sim \mathcal{N}(0, 1)$ with $\mathbb{E}(\epsilon|Y_!, Y_2) = 0$. It follows that

$$\begin{aligned}\widehat{Y}_0^N &= w_1 Y_1 + w_2 + Y_2 + \mu^* \\ &= \mu_0 + w_1(Y_1 - \mu_1) + w_2(Y_2 - \mu_2),\end{aligned}$$

where $w_1 = -q_1/q_0$ and $w_2 = -q_2/q_0$ and $\mu^* = \mu_0 - w_1\mu_1 - w_2\mu_2$. These results imply that there is no reason to impose the restrictions $w_1 \leq 0, w_2 \leq 0$ (positivity) and $w_1 + w_2 = 1$ (adding-up). Furthermore, the construction of SC should include a constant term, as otherwise the SC may have a different mean, See also [Doudchenko and Imbens, 2016] for a careful discussion of these restrictions.

For illustration assume that

$$\boldsymbol{y} \sim \mathcal{N}\left(\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1 & 0.1 & 0.4\\0.1 & 1 & 0.5\\0.4 & 0.5 & 1\end{pmatrix}\right)$$

Elaborate here once understood. For this example the optimal weights for the SC result as $w_1 = -0.133$, $w_2 = 0.4667$ and $\mu^* = 1 - w_1 - w_2 = 0.667$. Note that w_1 is negative even all bivariate correlations between the panel units are positive. One may argue that this solution does not make much sense as from a economic perspective it is not clear what it means that Y_1 enters the SC with a negative sign. This demonstrates the trade-off between optimality in a statistical sense and the interpretability of the solution.

What happens if we impose the restrictions that all weights are positive and sum up to unity? In this case the restricted optimum yields the linear combination $\widetilde{Y}_0^N = 0.2Y_1 + 0.8Y_2$. The important difference lies in the variance of these estimates. For our example we obtain

$$\begin{aligned}var(Y_0 - \widehat{Y}_0^N) &= 0.827\\ var(Y_0 - \widetilde{Y}_0^N) &= 1.160\end{aligned}$$

It is interesting to note that the variance of the restricted estimate is even larger than the unconditional variance of Y_0. This is possible as $(w_1, w_2) = (0, 0)$ is not included in the restricted parameter space.

It is not difficult to see that if Y_0 is not correlated with Y_1 and Y_2, then the optimal estimate boils down to $\widehat{Y}_0^N = \mu_0$ and therefore it does not make sense to involve a SC. In microeconometric settings it is usually assumed that the individuals in the treatment group and individuals in th econtrol group are uncorrelated. In such cases we do not care about constructing a SC. The crucial feature of SC methods is the correlation between the units in the treatment and the control group. In macroeconomic applications however, the variables in the treatment and control group (e.g. GDP) are typically correlated and it is therefore important to model the relationship between the variables. As the simple scenario with only two panel units in the donor pool is highly unrealistic in practice, we now move to the general static case with $k - 1$ panel units.

3.3. General Static Extension

What must be clear by now

- **Derive first analytical expressions for the case with k donors before talking about regularization**

In empirical practice it is often the case that the number of pre-intervention time periods $T_0 - 1$ is small and my even be smaller than the number of units in the donor pool, k.

3.4. General Dynamic Extension

What must be clear by now

- **TBD**

When modeling macroeconomic time series it is often assumed that the $(k + 1) \times 1$ vector of time series $y_t = (Y_{0t}, ..., Y_{kt})'$ can be represented by a VAR model given by

4. Simulation

- Simulation study is supposed to guide us to the full VARSC-model
- Ideas in OneNote

5. Applications

We consider three leading examples:

- [Abadie and Gardeazabal, 2003]
- [Abadie et al., 2010]
- [Abadie et al., 2015]

6. Conclusion

- Some concluding remark and an outlook
- Keep short, around 1-2 pages
- Natural extension: case with explanatory variables

References

[Abadie, 2021] Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425.

[Abadie et al., 2010] Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105:493–505.

[Abadie et al., 2015] Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

[Abadie and Gardeazabal, 2003] Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93:113–132.

[Abadie and Imbens, 2002] Abadie, A. and Imbens, G. (2002). Bias-corrected matching estimators for average treatment effects. *Journal of Business and Economic Statistics*, 29.

[Abadie and Imbens, 2006] Abadie, A. and Imbens, G. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74:235–267.

[Abadie and L'Hour, 2021] Abadie, A. and L'Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116:1–34.

[Amjad et al., 2018] Amjad, M., Shah, D., and Shen, D. (2018). Robust synthetic control. *Journal of the American Statistical Association*, pages 1–51.

[Andrews, 2003] Andrews, D. W. K. (2003). End-of-Sample Instability Tests. *Econometrica*, 71(6):1661–1694.

[Arkhangelsky et al., 2021] Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118.

[Athey et al., 2017] Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2017). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116.

[Athey and Imbens, 2016] Athey, S. and Imbens, G. (2016). The state of applied econometrics - causality and policy evaluation. *Journal of Economic Perspectives*, 31.

[Ben-Michael et al., 2021] Ben-Michael, E., Feller, A., and Rothstein, J. (2021). The augmented synthetic control method. *SSRN Electronic Journal*.

[Ben-Michael et al., 2021] Ben-Michael, E., Feller, A., and Rothstein, J. (2021). Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84.

[Born et al., 2019] Born, B., Müller, G. J., Schularick, M., and Sedláček, P. (2019). The Costs of Economic Nationalism: Evidence from the Brexit Experiment*. *The Economic Journal*, 129(623):2722–2744.

[Breitung and Knüppel, 2021] Breitung, J. and Knüppel, M. (2021). How far can we forecast? Statistical tests of the predictive content. *Journal of Applied Econometrics*, 36(4):369–392.

[Brodersen et al., 2015] Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2015). Inferring causal impact using bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1):247–274.

[Cattaneo et al., 2021] Cattaneo, M., Feng, Y., and Titiunik, R. (2021). Prediction intervals for synthetic control methods*. *Journal of the American Statistical Association*, 116:1–44.

[Chernozhukov et al., 2019] Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2019). Inference on average treatment effects in aggregate panel data settings. CeMMAP working papers CWP32/19, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

[Chernozhukov et al., 2021] Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021). An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls. University of California at San Diego, Economics Working Paper Series qt90m9d66s, Department of Economics, UC San Diego.

[Cho, 2020] Cho, S.-W. S. (2020). Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: The case of Sweden. *The Econometrics Journal*, 23(3):323–344.

[Cunningham, 2021] Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

[Doudchenko and Imbens, 2016] Doudchenko, N. and Imbens, G. W. (2016). Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis. NBER Working Papers 22791, National Bureau of Economic Research, Inc.

[Ferman, 2021] Ferman, B. (2021). On the Properties of the Synthetic Control Estimator with Many Periods and Many Controls. *Journal of the American Statistical Association*, 116(536):1764–1772.

[Firpo and Possebom, 2018] Firpo, S. and Possebom, V. (2018). Synthetic control method: Inference, sensitivity analysis and confidence sets. *Journal of Causal Inference*, 6(2).

[Frangakis and Rubin, 2002] Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1):21–29.

[Funke et al., 2020] Funke, M., Schularick, M., and Trebesch, C. (2020). Populist leaders and the economy. ECONtribute Discussion Papers Series 036, University of Bonn and University of Cologne, Germany.

[Hahn and Shi, 2017] Hahn, J. and Shi, R. (2017). Synthetic control and inference. *Econometrics*, 5(4).

[Hainmueller et al., 2011] Hainmueller, J., Diamond, A., and Abadie, A. (2011). Synth: An r package for synthetic control methods in comparative case studies. *Journal of Statistical Software*, 42.

[Hartford et al., 2017] Hartford, J., Lewis, G., Leyton-Brown, K., and Taddy, M. (2017). Deep iv: A flexible approach for counterfactual prediction. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 1414–1423. JMLR.org.

[Harvey and Thiele, 2020] Harvey, A. and Thiele, S. (2020). Cointegration and control: Assessing the impact of events using time series data. *Journal of Applied Econometrics*, 36.

[Kellogg et al., 2021] Kellogg, M., Mogstad, M., Pouliot, G. A., and Torgovitsky, A. (2021). Combining matching and synthetic control to tradeoff biases from extrapolation and interpolation. *Journal of the American Statistical Association*, 116(536):1804–1816. PMID: 35706442.

[Kuosmanen et al., 2021] Kuosmanen, T., Zhou, X., Eskelinen, J., and Malo, P. (2021). Design Flaw of the Synthetic Control Method. MPRA Paper 106328, University Library of Munich, Germany.

[Martin et al., 2012] Martin, V., Hurn, S., and Harris, D. (2012). *Econometric Modelling with Time Series: Specification, Estimation and Testing*. Themes in Modern Econometrics. Cambridge University Press.

[Muhlbach and Nielsen, 2019] Muhlbach, N. S. and Nielsen, M. S. (2019). Tree-based Synthetic Control Methods: Consequences of moving the US Embassy. Papers 1909.03968, arXiv.org.

[Neyman, 1923] Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science 5*, 4:465–472.

[Rosenbaum and Rubin, 1983] Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

[Rubin, 1974] Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

[von Brzeski et al., 2015] von Brzeski, V., Taddy, M., and Draper, D. (2015). Causal inference in repeated observational studies: A case study of ebay product releases.

Appendix

- Some proofs would be nice.